Postel

7127

Network Working Group
Request for Comments: 183
NIC: 7127

J. M. Winett
Lincoln Laboratory
21 July 1971

Categories: D.2, D.3
Obsoletes: None
Updates: None
Related: 109, 110, 105, 158

The EBCDIC Codes and Their Mapping to ASCII

Abstract:

To uniquely map the ASCII codes into corresponding EBCDIC codes in a consistent manner throughout the ARPA Network, this RFC describes and defines the IBM Standard Extended BCD Interchange Code.

Introduction:

The IBM Corporate Systems Standard, Extended BCD Interchange Code (EBCDIC) defines 8-bit graphic and control codes (See Figure 1). The basic EBCDIC code consists of 54 controls (including space) and 88 graphics. This set is extended to include 10 special graphics and 1 special control (EO). These special graphics originate from the 7-bit hollerith code and include 6 ASCII graphics. The EBCDIC code is further extended to include the publishing and printing graphics option which specifies 52 graphics. Of these graphics, 32 appear on the IBM TN print chain. Four of these graphics are duals with graphics not on the TN print chain, and one graphic (degree) is dual with a graphic in the special graphics set of the basic code (tilde).

It is desireable to uniquely map the ASCII codes into corresponding EBCDIC codes in a consistent manner throughout the ARPA network.

For each of the 34 ASCII controls (including space and delete) there is a corresponding EBCDIC control (assigning ASCII control DC3 to the EBCDIC code X'13'). For 85 of the 94 ASCII graphics, there is a corresponding graphic in the basic EBCDIC set. Three different correspondences can be made for the other 9 ASCII graphics.

This material has not been reviewed for public release and is intended only for use with the ARPA network. It should not be quoted or cited in any publication not related to the ARPA network.

I. IBM Correspondence

a) IBM recommends the following ASCII duals with basic EBCDIC graphics.

ASCII	EBCDIC	Code
[	¢	X'4A'
]	!	X'5A'
!	|	X'4F'
^	¬	X'5F'

Note that the EBCDIC graphic for exclamation point (!) is not chosen to correspond to the ASCII graphic for exclamation point (!), though this would be a sensible choice, and thus another code must be used to represent this graphic.

b) Special EBCDIC graphics would be used to represent the other ASCII graphics.

Graphic	Code
`	X'6A'
~	X'79'
\	X'A1'
{	X'E0'
}	X'C0'
	X'D0'

II. Publishing Correspondence

a) Associate the following special EBCDIC graphics with the corresponding ASCII graphics.

Graphic	Code
¦	X'6A'
¬	X'79'
~	X'A1'
\	X'E0'

b) Associate the following publishing EBCDIC graphics with the corresponding ASCII graphics.

Graphic	Code
^	X'71'
[	X'AD'
]	X'BD'
{	X'8B'
}	X'9B'

The codes for open bracket and close bracket are chosen since these graphics appear on the TN print chain. The codes for left brace and right brace are chosen rather than the codes in the special graphics set for opening brace and closing brace, respectively, since these graphics are similiar and also appear on the TN print chain.

III. Graphical Correspondence:

a) Associate the following basic EBCDIC graphics with the indicated ASCII graphics because of their graphic similiarity.

|	X'4F'	with	¦
¬	X'5F'	with	~

b) Associate the basic EBCDIC graphic for cent with the ASCII graphic for reverse slash.

¢	X'4A'	with	\

This choice is made since the cent graphic is not an ASCII graphic and is the only graphic in the basic EBCDIC set which would not otherwise be associated with any ASCII graphic.

c) Associate the special EBCDIC graphic grave accent.

` X'79'

with the corresponding ASCII graphic.

d) Associate the following publishing EBCDIC graphics with the corresponding ASCII graphics.

^	X'71'	@ SEX
[	X'AD'	B4
]	X'BD'	B5
{	X'8B'	B2
}	X'9B'	B3

The codes for open bracket and close bracket are chosen since these graphics appear on the TN print chain. The codes for left brace and right brace are chosen rather than the codes in the special graphics set for opening brace and closing brace, respectively, since these graphics are similiar and also appear on the TN print chain.

Standards:

In order that the mapping from ASCII into EBCDIC and vice versa could become standardized, I would appreciate comments on the above from each site whose operating system uses EBCDIC as the internal code.

Telnet Codes:

For those sites who may wish to provide or use TELNET services that communicate using an EBCDIC code, a standard code must be specified. The codes given in Figure 1 can form the basis for a standard. Specific codes must also be specified for the TELNET control codes. The following are suggested:

	Hex Code
Sync	38
break	39
NOP	3A
return to ASCII	FF
no echo	14
echo	23
hide input	24

To eliminate using one code for two graphics, I propose that the TN graphics be associated with their corresponding code. The graphic tilde (~) might be assigned to the code X'E1' rather than keeping the dual with the graphic for degree. This would have no effect if the Graphical Correspondence were chosen for the EBCDIC to ASCII mapping with the code X'5F' for logical not associated with tilde. The other graphics of the publishing and printing option (Double Acute, Inferior Hook, Macron, and Inferior Comma) which are not on the TN print chain but have the same codes as graphics on the TN print chain would not be considered to be part of the standard EBCDIC code.

EBCDIC Questionaire

1. For ASCII to EBCDIC mapping of the 9 special ASCII graphics do you prefer:

   a.) The IBM correspondence _____
   b.) The Publishing correspondence _____
   c.) The Graphical correspondence _____
   d.) Another correspondent (describe) _____

2. Do you concur with the definition of the standard EBCDIC code, including TELNET control codes?

   YES _____ NO _____

   Comments: _____

   _____

   _____

3. Please list for your operating system:

   a.) graphics not included in the complete EBCDIC code.
   b.) graphics given a different code.
   c.) controls given one of the graphic codes.
   d.) controls given one of the control codes but defined to be a different control.
   e.) all the controls which have meaning with your operating system (i.e., for which special action is taken) and state the action.

Reply from: Name _____

Telephone _____

Site _____

Host Computer _____

Send to:
Joel M. Winett
M.I.T. Lincoln Laboratory
Room C-151
Lexington, Mass. 02173

Or call:
(617) 862-5500 ext. 7474

	0 0	0 0	0 0	0 0	0 1	0 1	0 1	0 1	1 0	1 0	1 0	1 0	1 1	1 1	1 1	1 1
	1 0	0 0	0 0	0 0	1 1	1 1	1 1	1 1	0 0	0 0	0 0	0 0	1 1	1 1	1 1	1 1
	2 0	0 0	1 1	1 1	0 0	0 0	1 1	1 1	0 0	0 0	1 1	1 1	0 0	0 0	1 1	1 1
	3 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0	1 0
4567																
0000	NUL	DLE	DS		SP	&	-					-	0	{	}	\	0
0001	SOH	DC1	SOS				/		a	j		o	1	A	J		1
0010	STX	DC2	FS	SYN					b	k	s		2	B	K	S	2
0011	ETX	DC3 TM							c	l	t		3	C	L	T	3
0100	PF	RES	BYP	PN					d	m	u		4	D	M	U	4
0101	HT	NL	LF	RS					e	n	v		5	E	N	V	5
0110	LC	BS	ETB	UC					f	o	w		6	F	O	W	6
0111	DEL	IL	ESC	EOT					g	p	x		7	G	P	X	7
1000	GE	CAN							h	q	y		8	H	Q	Y	8
1001	RLF	EM							i	r	z		9	I	R	Z	9
1010	SMM	CC	SM		¢	!		:									
1011	VT	CU1	CU2	CU3	.	$	,	#									
1100	FF	IFS		DC4	<	*	%	@									
1101	CR	IGS	ENQ	NAK	(	)	_	'									
1110	SO	IRS	ACK		+	;	>	=									
1111	SI	IUS	BEL	SUB	|	¬	?	"								EO	

0	1	2	3	4	5	6	7

Code Structure

Figure 1.

Extended Binary-Coded Decimal Interchange Code (EBCDIC)

Publishing option graphics on the TN print chain

Publishing option graphics not on the TN print chain

Special EBCDIC graphics

8	0	0	0	0	0	0	0	0
7	0	0	0	0	1	1	1	1
6	0	0	1	1	0	0	1	1
5	0	1	0	1	0	1	0	1
4321								
0000	NUL	DLE	SP	0	@	P	`	p
0001	SOH	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BEL	ETB	'	7	G	W	g	w
1000	BS	CAN	(	8	H	X	h	x
1001	HT	EM	)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[	k	{
1100	FF	FS	,	<	L	\	l	|
1101	CR	GS	-	=	M	]	m	}
1110	SO	RS	.	>	N	^	n	~
1111	SI	US	/	?	O	_	o	DEL

8	7	6	5	4	3	2	1

Code Structure

Figure 2

USA Standard Code for Information Interchange (USASCII)

Hex Code	Category	Control	Name
00	CC	NUL	Null
01	CC	SOH	Start of Heading
02	CC	STX	Start of Text
03	CC	ETX	End of Text
04	DC	PF	Punch Off
05	FE	HT	Horizontal Tab
06	GR	LC	Lower Case
07	GR	DEL	Delete
08	GR	GE	Graphic Escape
09	FE	RLF	Reverse Line Feed
0A	CC	SMM	Start of Manual Message
0B	FE	VT	Vertical Tab
0C	FE	FF	Form Feed
0D	FE	CR	Carriage Return
0E	GR	SO	Shift Out
0F	GR	SI	Shift In
10	CC	DLE	Data Link Escape
11	DC	DC1	Device Control 1
12	DC	DC2	Device Control 2
13	DC	TM/DC3	Tape Mark /Device Control 3
14	DC	RES	Restore
15	FE	NL	New Line
16	FE	BS	Backspace
17	DC	IL	Idle
18	GR	CAN	Cancel
19	DC	EM	End of Medium
1A	DC	CC	Curser Control
1B	CU	CU1	Customer Use 1
1C	IS	IFS	Info. Field Separator
1D	IS	IGS	Info. Group Separator
1E	IS	IRS	Info. Record Separator
1F	IS	IUS	Info. Unit Separator

Figure 3: EBCDIC Control Functions

Hex Code	Category	Control	Name
20	ED	DS	Digit Select
21	ED	SOS	Start of Significance
22	ED	FS	Field Separator
23			(Reserved)
24	DC	BYP	Bypass
25	FE	LF	Line Feed
26	CC	ETB	End of Text Block
27	GR	ESC	Escape
28			(Reserved)
29			(Reserved)
2A	DC	SM	Set Mode
2B	CU	CU2	Customer Use 2
2C			(Reserved)
2D	CC	ENQ	Enquiry
2E	CC	ACK	Acknowledge
2F	DC	BEL	Bell
30			(Reserved)
31			(Reserved)
32	CC	SYN	Synchronous Idle
33			(Reserved)
34	DC	PN	Punch On
35	DC	RS	Reader Stop
36	GR	UC	Upper Case
37	CC	EOT	End of Transmission
38			(Reserved)
39			(Reserved)
3A			(Reserved)
3B	CU	CU3	Customer Use 3
3C	DC	DC4	Device Control 4
3D	CC	NAK	Negative Acknowledge
3E			(Reserved)
3F	GR	SUB	Substitute

Figure 3: EBCDIC Control Functions
(Continued)

CC (Communication Control). A functional character intended to control or facilitate transmission of information over communication networks.

FE (format Effector). A functional character which controls the layout or positioning or information in printing or display devices.

IS (Information Separator). A character which is used to separate and qualify information in a logical sense. There is a group of four such characters, which are to be used in a hierarchical order.

DC (Device Control). A functional character used for the control of ancillary devices associated with data processing or telecommunication systems, more especially switching devices "on" and "off."

ED (Edit and Mark). A control character used by the System/360 Edit and Mark (EDMK) instruction for the formatting of alphanumeric fields.

GR (Graphic Control). A control character indicating that the code combinations which follow are to be interpreted in a particular code table, depending upon the particular control character.

CU (Customer Use). A character excluded from future assignment by IBM. These "protected" codes are intended for use by customer systems so that their use will not conflict with a possible future IBM use.

Figure 4

Categories of Control Functions

	Hex Code	Graphic	Name
*	6A	¦	Vertical Line
*	79	`	Grave Accent
*	A1	~	Tilde
*	C0	{	Opening Brace
	CC	⌠	Hook
	CE	╪	Fork
*	D0	}	Closing Brace
*	E0	\	Reverse Slant
	EC	┤	Chair
	FA	˙	Long Vertical Mark
	FF	EO	Eight Ones

Figure 5: Special EBCDIC Graphics

*ASCII Graphic

	Hex Code	Graphic	Name
**	A0	-	Superscript Minus
*	A1	°	Degree
*	B0	⁰	Superscript Zero
*	B1	¹	Superscript One
*	B2	²	Superscript Two
*	B3	³	Superscript Three
	B4	⁴	Superscript Four
	B5	⁵	Superscript Five
	B6	⁶	Superscript Six
	B7	⁷	Superscript Seven
	B8	⁸	Superscript Eight
	B9	⁹	Superscript Nine
	8B	{	Left Brace
	8C	≤	Equal or Less Than
	8D	⁽	Superscript Left Parenthesis
	8E	⁺	Superscript Plus Sign
	8F	⌖	Plotting Cross
	9B	}	Right Brace
	9C	⧫	Lozenge
	9D	⁾	Superscript Right Parenthesis
	9E	±	Plus or Minus
	9F	▯	Histogram
	AB	⌊	Lower Left Corner
	AC	⌜	Upper Left Corner
	AD	[	Open Square Bracket
	AE	≥	Equal or Greater Than
	AF	∘	Bullet (Plotting Circle)
	EB	⌋	Lower Right Corner
	EC	⌝	Upper Right Corner
	ED	]	Close Square Bracket
	BE	≠	Not Equal
	BF	—	Extended Dash

Figure 6: Publishing and Printing Graphics
also on the TN Print Chain

** Dual with the special EBCDIC graph c tilde

* Dual with another graphic which is not on the TN print chain

	Hex Code	Graphic	Name
	70	°	Scandanavian Accent
	71	^	Circumflex
	72	¨	Diaeresis
	73	/	Diacritical Virgule
	74	´	Acute Accent
	75	˙	Superior Dot
	76	¸	Cedilla
	77	˘	Breve
	78	ˇ	Caron
	8A	↑	Up Arrow
	9A	†	Dagger
*	B0	˝	Double Acute
*	B1	˛	Inferior Hook
*	B2	¯	Macron
*	B3	¸	Inferior Comma
	CD	‘	Open Quote
	DB	£	Pound Sign
	DC	§	Section Sign
	DD	¶	Paragraph Sign
	ED	’	Close Quote

Figure 7: Publishing and Printing Graphics not on the TN Print Chain

* Dual with another graphic which is on the TN print chain

Name	Graphic	Hex Code	Graphic	Name
Tilde	~	A1	°	Degree
Double Acute	˝	B0	⁰	Superscript Zero
Inferior Hook	˛	B1	¹	Superscript One
Macron	¯	B2	²	Superscript Two
Inferior Comma	¸	B3	³	Superscript Three

Figure 8: Graphic Duals

Codes	Graphics	Names
AF--75	·	Bullet--Superior Dot
8B--C0	{	Left Brace--Opening Brace
9B--D0	}	Right Brace--Closing Brace
61--73	/	Slash--Diacritical Virgule
A1--70	°	Degree--Scandanavian Accent
4F--FA	|	Logical Or--Long Vertical Mark
6B--76--B3	,	Comma--Cedilla--Inferior Comma
60--B2	-	Dash--Macron

Figure 9: Similiar Graphics

Name	Control	Hex Code	Graphic	Name
Group Mark	GM	4F	|	Logical Or
Mode Change	MC	5F	¬	Logical Not
Plus Zero	PZ	C0	{	Opening Brace
Minus Zero	MZ	D0	}	Closing Brace
Record Mark	RM	E0	\	Reverse Slant
Word Separator	WS	6D	–	Underscore
Segment Mark	SM	6F	?	Question Mark
Substitute Blank	SB	7A	:	Colon
Tape Mark	TM	7F	"	Quotation Marks

Figure 10: Graphic and Control Duals